Sample Size

- The moment of truth
- Where the rubber hits the pavement

Hermetic Decalogue

… Thou shalt not sit

With statisticians…

WH Auden, 1949±2

QoL in Trial of Brain Metastases

- Endpoint: FACT-Br
  - Mean = 50
  - SD = 10

The raw ingredients

- What is your question, precisely?
- What is your outcome, precisely?
- Who will be measured?
- Type 1 and Type 2 error rates
- The variability

Difference to be detected

- Clinically important?
- Biologically credible?
- The "I would kick myself" difference

Does time play a role?

- Pattern of recruitment
- Follow-up time
- Hazard over time
- Hazard ratio over time
- Competing risks

QoL Trial

- Primary tumor may be
  - Lung
  - Breast
  - Other
- Brain mets may be anywhere – disability depends on size and location
- Drug designed to shrink mets

Operating characteristics

- Type 1 error rate = 0.05 two sided
- Type 2 error rate = 0.90

What is the question, precisely?

- Does the drug improve QoL?
- Does the drug improve outcome on the FACT-Br?

What is the…

- Mean difference in FACT-Br?
  - Variability: SD=10
- Difference in proportions falling below 30?
  - Variability is binomial
- "Difference" in time to falling below 30?
  - Variability: hazard and hazard ratio

Generic Formula

Sample size *per group*:

$$\frac{2\sigma^2 (z_1+z_2)^2}{\delta^2}$$

$$(z_1+z_2)^2$$

$$\frac{2\sigma^2 (z_1+z_2)^2}{\delta^2}$$

- A fixed number
- You choose!
- If 0.05 and 0.90, this quantity is about 10
- Kick-yourself power: β=.5 and z_2=0.
- Subscripts usually: (1-α)/2 and (1-β)

δ^2

$$\frac{2\sigma^2 (z_1+z_2)^2}{\delta^2}$$

The difference you _____ detect.

a) want to

b) believe is clinically meaningful

c) believe is biologically credible

d) can afford to

σ^2

- From:
  - Past data
  - Assumptions in study
- Very often underestimated!
  - Past data not directly relevant
  - Problems in study inflate the variance

δ^2

$$\frac{c}{2\sigma^2 (z_1+z_2)^2}$$

$$\delta^2$$

The difference you _____ detect.

a) want to

b) believe is clinically meaningful

c) believe is biologically credible

d) can afford to

2

$$\frac{2\sigma^2 (z_1+z_2)^2}{\delta^2}$$

- **The 2 is *per group***
- **The factor for a two-group study is 4.**

$$\mathrm{Var}(\bar{x}-\bar{y}) = \mathrm{Var}(\bar{x}) + \mathrm{Var}(\bar{y})$$

$$= 2\mathrm{Var}(\bar{x}) = 2\sigma^2/n$$

Case #1: Recruitment and follow-up

- Everyone is recruited at the same time
- No one dies or is lost to follow-up
- Everyone is followed for exactly 1 year

Endpoint: difference in mean

$$\frac{2\sigma^2 (z_1+z_2)^2}{\delta^2}$$

- Assume the mean is normal

  $\sigma=10$; $\delta=10$

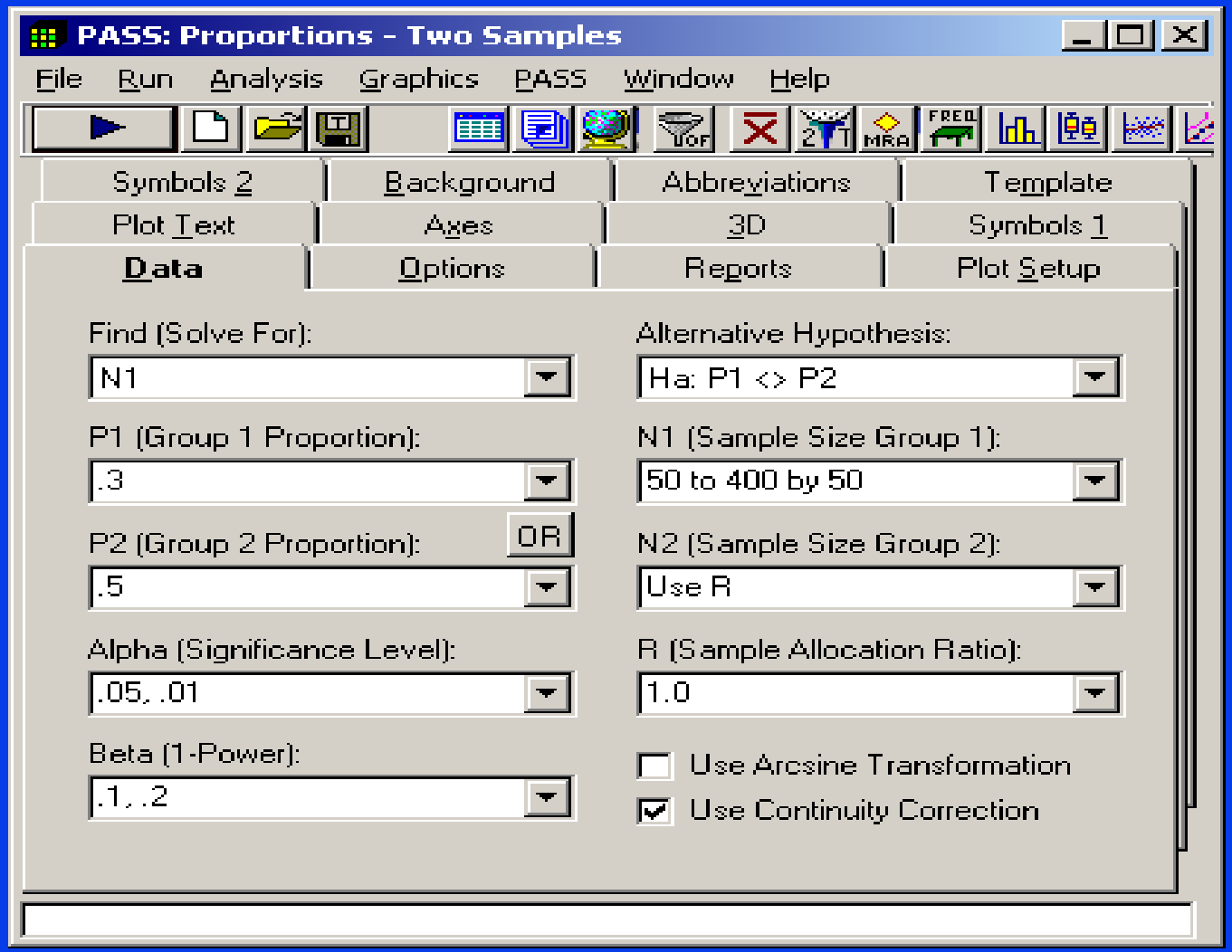
- Sample size = 2 (100)(10)/100 = 20/group
- Doubling the SD or halving the difference quadruples the sample size

Endpoint: proportion falling below 30

- (Proportion falling at least 10 points)
- (Proportion falling at least 20 percent)
- Say we want to compare 50 percent vs. 30 percent:

$$\frac{2(\text{binomial variance})^2 (z_1+z_2)^2}{\delta^2}$$

e.g., PASS


PASS: Proportions - Two Samples
File Run Analysis Graphics PASS Window Help
Symbols 2 Background Abbreviations Template
Plot Text Axes 3D Symbols 1
Data Options Reports Plot Setup
Find (Solve For):
N1
Alternative Hypothesis:
Ha: P1 <> P2
P1 (Group 1 Proportion):
.3
N1 (Sample Size Group 1):
50 to 400 by 50
P2 (Group 2 Proportion):
OR
.5
N2 (Sample Size Group 2):
Use R
Alpha (Significance Level):
.05, .01
R (Sample Allocation Ratio):
1.0
Beta (1-Power):
.1, .2
Use Arcsine Transformation
Use Continuity Correction

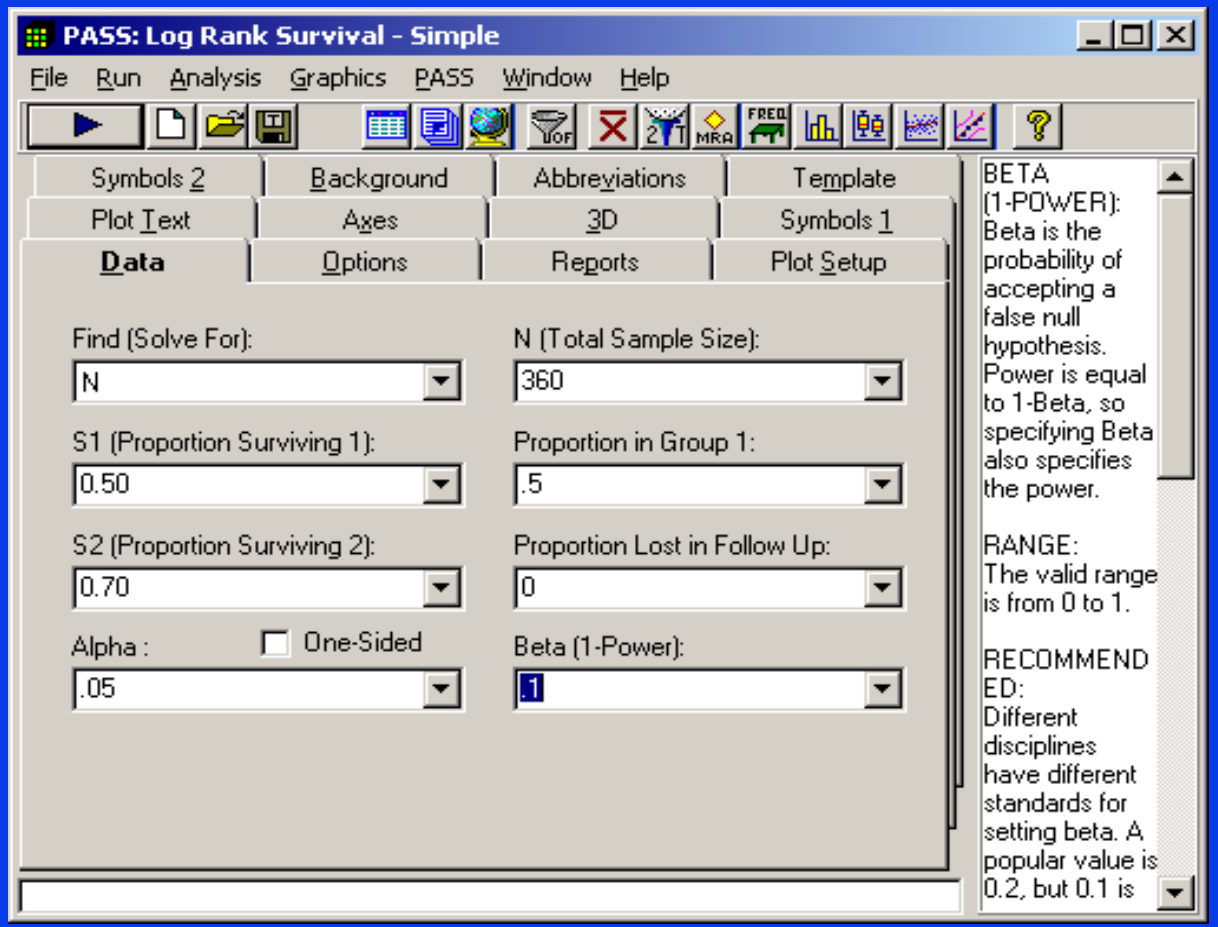
Binomial answer

- 130 per group
- If only 80 percent power, 100 per group
- If Type 1 error rate is 0.01 and power =
  - 90% n per group = 185
  - 80% n per group = 150

Time to falling below 30

- Assume exponential time to failure
- Assume that at 4 months 50% of control and 70% of treatment are still above 30
- Required sample size is 128 per group.

PASS


PASS: Log Rank Survival - Simple
File Run Analysis Graphics PASS Window Help
Symbols 2 Background Abbreviations Template
Plot Text Axes 3D Symbols 1
Data Options Reports Plot Setup
Find (Solve For):
N
N (Total Sample Size):
360
S1 (Proportion Surviving 1):
0.50
Proportion in Group 1:
.5
S2 (Proportion Surviving 2):
0.70
Proportion Lost in Follow Up:
0
Alpha :
One-Sided
.05
Beta (1-Power):
.1
BETA (1-POWER):
Beta is the probability of accepting a false null hypothesis. Power is equal to 1-Beta, so specifying Beta also specifies the power.
RANGE:
The valid range is from 0 to 1.
RECOMMENDED:
Different disciplines have different standards for setting beta. A popular value is 0.2, but 0.1 is

Minor headaches

- Distribution of the mean not normal
- Population heterogeneous

Major headaches

- Missing data
- Time of follow-up (all three, but problem explicit in time-to-failure)
- Non-exponential failure
- Non-proportional hazards

Missing data

- ….more next week
- For now:
  - Common approaches

    Just Ignore

    Last Observation Carried Forward (LOCF)

    Something more complicated
  - My principle: you should not win because of missing data

Implications for Sample Size

- You need 100/group and expect 10% missing
- LOCF people would say: 100
- Just Ignore people would say: 111
- Lavori's rule of thumb: each missing person = 3 observed

Therefore, your sample size should be
90+3(10)=120

Time: Exponential/non-exponential

- Light bulb model often works well
- All we need to know is person-years of follow-up
- So, 4 people followed 1 year = 1 person followed 4 years

Recruitment: exponential case

- Follow-each person 12 months-recruitment pattern doesn't matter for sample size
- Follow each person until the last recruited has 12 months of follow-up
  - Persons years of follow-up depends on recruitment patter
  - The SLOWER the recruitment, the SMALLER the sample size

Non-exponential examples

- Post-CABG surgery:
  - Cognition impaired at first perhaps as consequence of anesthesia
  - Long-term may show slight decline, perhaps consequence of mini-strokes

Non-proportional hazards

- Landmark vs. log-rank time to failure
- E.g., time to diabetes
  - Control
  - Diet
  - Drug
- If we stop at two years, we have no data for four years

Moral

- Don’t do sample size calculation in a rush
- Use standard software to help but the big problem is not the calculation, it’s gathering the raw materials